



**You have downloaded a document from
RE-BUS
repository of the University of Silesia in Katowice**

Title: Параллельный корпус InterCorp в переводной лексикографии

Author: Andrzej Charciarek

Citation style: Charciarek Andrzej. (2017). Параллельный корпус InterCorp в переводной лексикографии. "Opera Slavica (Brno)" (2017), nr 2, s. 5-17.



Uznanie autorstwa - Na tych samych warunkach - Licencja ta pozwala na kopiowanie, zmienianie, rozprowadzanie, przedstawianie i wykonywanie utworu tak długo, jak tylko na utwory zależne będzie udzielana taka sama licencja.



UNIWERSYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

Параллельный корпус InterCorp в переводной лексикографии

Parallel Corpus InterCorp in Translation Lexicography

Анджей Харциарек

(Катовице, Польша)

Абстракт:

Настоящая статья посвящена описанию возможностей, которые параллельные корпуса предоставляют переводной лексикографии. Многие годы параллельные корпуса отдельных модулей национальных корпусов содержали преимущественно художественные тексты и их переводы. Со временем эта ситуация стала меняться. Лучший пример этому — корпус InterCorp, составляющий часть Чешского национального корпуса (ČNK). Пополнение его текстами различных жанров создало совершенно новые перспективы для лексикографических исследований, в том числе и для переводной лексикографии. Это источник, который позволяет увеличить количество возможных переводных эквивалентов, определить значение и частотность употребления лексических единиц в отдельных жанрах, выявить идиомы и их варианты, верифицировать лексикографическую информацию в уже существующих словарях и т. п. В статье приводятся конкретные примеры использования параллельных корпусов в переводной польско-русской лексикографии.

Ключевые слова:

переводная лексикография; перевод; параллельный корпус InterCorp; польский язык; русский язык

Abstract:

This article presents possibilities in which parallel corpora can be applied in translation lexicography. In the past the parallel corpora mostly included literary texts and their translations but this has started to change. It is best illustrated by the InterCorp multilingual corpus which is a part of the Czech National Corpus (ČNK) also including other genres. The appearance of the new genres in the corpus provides new research perspectives on lexicography and translation lexicography in particular. Due to them it is possible to increase the number of translation equivalents, determine the meaning and usage rate of specific lexical units in particular genres, identify idioms and their variations or verify information provided in the existing dictionaries etc. The article also provides specific examples illustrating how parallel corpora can be used in Polish-Russian translation lexicography.

Key words:

translation lexicography; translation; parallel corpus InterCorp; Polish language; Russian language

Возникновение корпусной лингвистики как самостоятельной научной дисциплины изменило подход не только к формам исследования языка, но и к принципам составления словарей. Массивы текстов и новые методы их обработки в возникающих лингвистических корпусах открыли принципиально новые возможности для одноязычной и переводной лексикографии. Национальные корпуса польского, русского и чешского языков, находящиеся в центре нашего внимания, постоянно расширяют свои рамки. В их составе возникают новые подкорпусы, а также внедряются новые инструменты для разного вида анализа языкового материала. Благодаря систематическому пополнению новыми текстами различных жанров лингвистические корпуса всё объективнее отражают языковую систему, предоставляя, возможно, наиболее достоверную информацию о её состоянии. Все эти факты имеют существенное значение для лексикографов, ищущих оптимальные корпуса текстов для составления словарей, в том числе переводных.

Составление словаря на основе лингвистического корпуса — идея не новая. Пример этому — проект лингвистов-лексикографов из Института польского языка ПАН, которые именно таким образом создают *Большой словарь польского языка*. Совершенно очевидно, что процесс создания подобного одноязычного словаря сильно отличается от составления переводного словаря. Задача последнего заключается в том, чтобы предоставить пользователю словаря соответствующие эквиваленты, необходимые в процессе перевода.

Как правило, эквиваленты понимаются исследователями двояко: как словарные (системные) и как текстовые (функциональные). Названия эквивалентов отражают виды эквивалентности, в первом случае — словарной, во втором — текстовой. В случае фразем, т.е. сверхлексемных единиц, нас интересует второй вид. То, что некоторые фраземы не фиксируются ни в общих, ни во фразеологических словарях, стало уже традицией. Главная причина такого положения вещей заключается в трудности определения отношения эквивалентности между фраземой в исходном языке (трансляндом) и соответствующей ей фраземой в языке перевода (транслятом). Так, затруднения в составлении пары эквивалентов могут возникнуть на любом этапе этого процесса — в определении значения транслянда, в определении значения транслята и, наконец, в определении их эквивалентности. Условием установления равнозначности транслянда и транслята является единственный критерий — ситуация употребления. Именно она становится основой перевода — *tertium comparationis*. Критерий ситуативности как ключевой отмечается и исследователями. В. Хлебда, ссылаясь на идеи А. Богуславского, предлагает процедуру «прагматической рамки», которая по сути своей учитывает все компоненты ситуации общения¹:

[КТО] — [КОГДА] — [ГДЕ] — [С КАКОЙ ЦЕЛЬЮ] — [О ЧЁМ] — [КАКИМ СПОСОБОМ]
ОБРАЩАЕТСЯ
[К КОМУ]
[ПО КАКОЙ ПРИЧИНЕ] [С КАКИМ РЕЗУЛЬТАТОМ]

Таким образом, критериев выбора оптимального переводного эквивалента несколько. В принципе о полной эквивалентности, если таковая вообще существует, можно говорить только тогда, когда переводной эквивалент удовлетворяет всем вышеназванным критериям.

Установление пары эквивалентов, по А. Богуславскому, возможно в случае достижения максимальной ситуативной адекватности, которая осуществляется самым репрезентативным, самым компетентным и самым информированным носителем обоих языков². Учёный констатирует, что транслят должен обладать максимумом функциональных характеристик транслянда. Этот максимум определяется в отношении к эмпирическим ситуативным данным с учётом дополнительных принципов:

- 1 CHLEBDA, W.: „Ramka pragmatyczna” w procesie weryfikowania i tworzenia słownikowych definicji frazeologizmów. In: HORDY, M. – MOKIJENKO, W. – WALTER, H. (red.): Język. Człowiek. Dyskurs. Szczecin 2007, s. 185–195.
- 2 BOGUSŁAWSKI, A.: Podstawy konfrontatywnej lingwistyki przekładowej. Łask 2013, s. 64.

- принципа системной конвенциональности функциональных различий исходного текста и текста перевода,
- принципа приоритета познавательных ценностей,
- принципа соизмеримой с исходным текстом стандартности текста перевода,
- принципа максимального сходства исходного текста и текста перевода в силу дифференциации трансландов и транслатов и
- принципа обратного перевода³.

Как можно видеть, установление пары эквивалентов представляется процессом весьма непростым, требующим от лексикографов, что часто постулировалось и постулируется, лингвистического анализа больших массивов текстов. Такие возможности, безусловно, открывают национальные корпуса языков. Их использование как в одноязычной, так и переводной лексикографии постепенно расширяется. Об этом свидетельствуют, по крайней мере, следующие факты:

- постоянно увеличивается объем корпусных текстов,
- повышается достоверность статистических данных,
- совершенствуются технические инструменты обработки текстов.

Всё бóльшим потенциалом обладают интересующие нас в данной статье параллельные корпуса, которые представляют для переводной лексикографии особую ценность.

Как известно, это источник, с помощью которого можно:

- выявлять и отрабатывать переводческие приёмы и лексико-грамматические трансформации;
- верифицировать лексикографическую информацию в уже существующих словарях;
- увеличивать число возможных переводных эквивалентов;
- определять частотность употребления языковых единиц в отдельных жанрах;
- отслеживать и фиксировать устойчивые словосочетания и идиоматические выражения.

Следует отметить, что многие годы содержание параллельных корпусов, отдельных модулей национальных корпусов, составляли художественные тексты и их переводы. Этот факт вполне естественен, поскольку качество языка художественной литературы и её переводов удовлетворяет самым строгим требованиям. Данное утверждение относится преимущественно к классической литературе, которая, в отличие от современной, гораздо сильнее влияла на формирование языковой нормы. Следовательно, роль современной художественной литературы в общественно-культурной жизни, учитывая

3 BOGUSŁAWSKI, A.: *Podstawy konfrontatywnej lingwistyki przekładowej*. Łask 2013, s. 64–79.

возможность её влияния на формирование языковой компетенции членов общества, сильно ограничена. Несмотря на её элитарный или массовый характер, современная литература далеко не в такой степени, как раньше, формирует наши языковые навыки и речевые умения. Следовательно, параллельные корпуса должны пополняться текстами различных авторов, стилей, жанров и, что самое важное, значительно увеличивать свои, как правило, небольшие объёмы.

Это хорошо понимают и корпусные лингвисты, которыми этот вопрос постепенно решается. Пример этому — параллельный корпус InterCorp v8 Национального корпуса чешского языка (ČNK), содержащий, кроме художественных текстов, так называемые коллекции (kolekce), среди которых находятся:

- публицистические статьи и выпуски новостей на сайтах Project Syndicate и VoxEurop,
- юридические тексты Евросоюза в корпусе Acquis Communautaire,
- записи заседаний парламента Евросоюза на период 2007–2011 гг. в корпусе Europarl
- титры на сайте OpenSubtitles.

Правда, вышеприведённые коллекции в такой же степени не касаются всех языков, напр. InterCorp v8, по очевидным причинам, не содержит текстов на русском языке из корпусов Europarl, Acquis Communautaire и VoxEurop. Польский InterCorp v8 — один из самых больших по величине корпусов объёмом в 80 млн. слов — содержит все вышеуказанные коллекции, за исключением Project Syndicate. Любой иноязычный текст параллельного корпуса имеет чешскую версию, поэтому по объёму чешский InterCorp v8 с другими корпусами сравнивать нельзя. Принципиально важно то, что добавление в параллельный корпус InterCorp v8 записей заседаний парламента Евросоюза и субтитров увеличило в нём долю устного материала.

Пополнение корпуса InterCorp v8 текстами других жанров создало совершенно новые перспективы для лексикографических исследований. Как уже было отмечено, в прошлом параллельные корпуса опирались в основном на художественные произведения и их переводы. Это был естественный способ достаточно быстрого создания такого вида корпуса, который одновременно гарантировал, как правило, высокое качество перевода. Тем самым текст перевода становился источником переводных эквивалентов. Казалось бы, такая модель идеальна с точки зрения перевода. Однако пользователи параллельных корпусов довольно быстро заметили их существенные недостатки. Они касаются прежде всего самой природы языка художественной литературы, который по сути своей должен отличаться от языка повседневного общения: быть оригинальным, необычным, изысканным. Следовательно,

язык художественного текста мог содержать и по существу содержал, напр., малочастотные коллокации, нераспространённые идиомы или авторские неологизмы. Очевидно, что многие предложения могут быть построены грамматически правильно, отражать возможности языковой системы, но одновременно входить в противоречие с узусом — носитель данного языка сразу скажет: «так не говорят». Здесь опять-таки затрагивается вопрос прагматики, поскольку именно к ней относится приведённое выражение «так не говорят». Практическое отсутствие устных текстов в корпусах было основным препятствием для описания языковой системы в употреблении, т. е. в речи. Их отсутствие можно объяснить тем, что отражение речи в письменной форме требует особых операций, которые гораздо сложнее составления корпуса изначально письменных текстов. Однако острая нехватка фиксации устной речи, учитывающей не столько публичную, сколько всего непубличную речь, приводит к созданию неполной картины языковой системы. Попутно отметим, что это и традиционная слабая сторона переводных словарей, которые в далеко недостаточной степени отражают устную непубличную речь. Это преимущественно реальная непринуждённая, спонтанная речь в условиях прямого контакта коммуникантов в неофициальной сфере общения, напр., непринуждённый разговор, а также разговор по телефону, рассказ, пересказ и т. д. Г. А. Китайгородская и Н. Н. Розанова выделяют виды непринуждённого разговора: домашний разговор, микродиалоги (в больнице, на почте, в аптеке, в библиотеке, в доме отдыха, в кассе, в магазине, в лифте, в транспорте, дома, информационные, на кухне, на пороге, на работе и т. п.), праздный разговор, разговор-воспоминание, разговор (в магазине, в парикмахерской, при встрече и т. п.), спор, телефонный разговор⁴. Все вышеприведённые виды должны быть в известной степени представлены в корпусе. Кроме того, следует и учесть кинотранскрипты, которые, напр., составляют отдельный подкорпус в устном корпусе НКРЯ. С. О. Савчук и Е. А. Гришина отмечают, что речь кино следует рассматривать как точную имитацию устной речи, которая по отдельным показателям (напр., дискурсивным маркерам) ближе либо к публичной, либо к непубличной речи⁵. Таким образом, речь кино нельзя не признать богатым источником лексикографического материала.

Как уже мы отметили, среди текстов, включённых в параллельный корпус InterCorp v8Я, нашлись и субтитры. Они, на наш взгляд, представляют особую ценность для лексикографов. Они отражают не только разные сферы

4 KITAJGORODSKAJA, M. V. – ROZANOVA, N. N.: *Reč' moskvičej. Kommunikativno-kul'turologičeskij aspekt*. Moskva 1999, s. 225.

5 GRIŠINA, Je. A. – SAVČUK, S. O.: *Korpus ustnyh tekstov v NKRJA: sostav i struktura*. In: Nacional'nyj korpus russkogo jazyka: 2006–2008. Novyje rezul'taty i perspektivy. Sankt-Peterburg 2009, s. 144.

функционирования устной речи, но и разные типы текстов, в том числе те, которые кажутся наиболее ценными: беседу, дискуссию и разговор. Именно речевых фраз, которые выражают различные эмоциональные состояния коммуникантов, явно не хватает в переводных словарях. Проблема фиксации подобных выражений заключается в трудности сформулировать их словарную дефиницию. На наш взгляд, она должна приобрести прагматический характер, т. е. учитывать прежде всего прагматические аспекты высказывания. Именно поэтому корпус InterCorp v8, содержащий субтитры, может в значительной мере решить эту проблему.

Яркий пример этому польская фраза *mam tego dość*, которая чаще всего не фиксируется переводными или фразеологическими словарями, и для которой попытаемся установить эквивалент/эквиваленты в русском языке. В *Ином словаре польского языка* под редакцией М. Банько она определяется следующим образом:

„Jeżeli ktoś ma czegoś lub kogoś **dość**, to chce, żeby się to skończyło lub żeby dana osoba przestała z nim być lub coś robić”⁶.

Параллельный корпус польско-русский InterCorp v8 фиксирует 23 вхождения этой фразы:

InterCorp v8 - Polish		InterCorp v8 - Russian	
._SUBTITLES	Już mam tego dość!	._SUBTITLES	С меня хватит, я не клону на эту ложь!
._SUBTITLES	Mam tego dość.	._SUBTITLES	С меня хватит.
._SUBTITLES	Mam tego dość.	._SUBTITLES	Я чертовски устал от всего этого!
._SUBTITLES	I mam tego dość.	._SUBTITLES	Меня шатает ветром.
Hasek-OsudyDobrehoVvSV	<p> - Mam tego dość! - wrzasnął frajter. </p>	Hasek-OsudyDobrehoVvSV	<p> - Хватит с меня! - крикнул вдруг ефрейтор. </p>
._SUBTITLES	Mam tego dość, kończymy z tym.	._SUBTITLES	Знаешь что, давай кончай с этим дерьмом.
._SUBTITLES	Mam tego dość!	._SUBTITLES	- Послушай ... - Не нужно!
._SUBTITLES	Przysięgam na Boga, mam tego dość.	._SUBTITLES	Блин, ей-богу, уже одышка началась.
._SUBTITLES	Mam tego dość.	._SUBTITLES	Меня уже тошнит от этого.
._SUBTITLES	Mam tego dość.	._SUBTITLES	Как же я устала от вас двоих!
._SUBTITLES	Mam tego dość.	._SUBTITLES	А мне надоело угождать тебе все время.
._SUBTITLES	Mam tego dość.	._SUBTITLES	Мне надоела эта мерзость.
._SUBTITLES	- Mam tego dość.	._SUBTITLES	— С меня хватит. — Хорошо.
._SUBTITLES	- Mam tego dość, do chuja pana.	._SUBTITLES	- Меня уже тошнит от тебя.
._SUBTITLES	Mam tego dość!	._SUBTITLES	Мне надоело от нее получать!
._SUBTITLES	Mam tego dość.	._SUBTITLES	Мне всё это тоже не в радость, Скотт.
._SUBTITLES	Mam tego dość!	._SUBTITLES	Мне это не нужно.
Hasek-OsudyDobrehoVvSV	<p> - Już mam tego dość.	Hasek-OsudyDobrehoVvSV	Теперь в постель и дрыхни! Понял? </p>
._SUBTITLES	Mam tego dość.	._SUBTITLES	- Кажется четваро.
._SUBTITLES	I mam tego dość.	._SUBTITLES	Я уже сказал - два автобуса и самолет. Ага, как же! И места в ложе на стадионе Янки!
._SUBTITLES	Mam tego dość, zdecyduj się na coś.	._SUBTITLES	Мне надоели твои перемены настроения.
._SUBTITLES	Mam tego dość!	._SUBTITLES	Всему есть предел!
._SUBTITLES	- Mam tego dość!	._SUBTITLES	- Я не могу так играть.

Само собой разумеется, что автоматическое выравнивание, а такое проведено в данном корпусе, нередко приводит к неточностям, и даже грубым ошибкам. Это и подтверждают некоторые найденные вхождения, которые не имеют ника-

6 BAŃKO, M. (red.): *Inny słownik języka polskiego*. T. 1. Warszawa 2000, s. 308.

кого отношения к возможным переводным эквивалентам. Так, пользователь корпуса InterCorp v8 должен тщательно проверить все вхождения. Понятно, что устранение сомнительных эквивалентов требует от него языковой компетенции. Несмотря на это большинство вхождений представляет собой ценный материал для лексикографической обработки.

Собранный корпусный материал показывает, что польско-русскую пару эквивалентов составляют фраземы *tam tego dość* и *с меня хватит*. Однако, следует обратить внимание и на вариант *хватит с меня*. Это лишь одно вхождение и поэтому надо обратиться к Национальному корпусу русского языка (далее — НКРЯ) с целью проверить частотность обоих вариантов. Статистика такова: *с меня хватит* — 103 вхождения в 84 документах; *хватит с меня* — 115 вхождений в 100 документах.

Оказывается, что оба варианта почти одинаково употребляются русскими и, следовательно, оба должны быть учтены в переводном словаре.

Кроме того, правая сторона параллельного корпуса приводит и другие возможные переводные эквиваленты, требующие подробной проверки.

Параллельный корпус русско-польский InterCorp v8 даёт следующие результаты *всему есть предел*:

InterCorp v8 - Russian		InterCorp v8 - Polish	
<input type="checkbox"/> rowlingova-hpot_tajemn	Нуж дудки, всему есть предел .	rowlingova-hpot_tajemn	0 nie, mam już tego dość, idę do Dumbledora! ... </p>
<input type="checkbox"/> _SUBTITLES	Но, конечно, всему есть предел .	_SUBTITLES	Ale oczywiście, są granice.
<input type="checkbox"/> _SUBTITLES	Моя жена была очень терпелива, но всему есть предел .	_SUBTITLES	Moja żona była bardzo cierpliwa jednak wszystko ma swoje granice.
<input type="checkbox"/> _SUBTITLES	Всему есть предел !	_SUBTITLES	Mam tego dość!
<input type="checkbox"/> rowlingova-hpot_vizen	Но всему есть предел !	rowlingova-hpot_vizen	<p> - Jestem zachwycona z wybranej Gryffindoru, ale to już przestaje być zabawne!

Как можно видеть, в определённых контекстах и *всему есть предел* может быть переводным эквивалентом *tam tego dość*. Однако малое количество вхождений (5) требует дополнительной проверки употребления *всему есть предел* в НКРЯ, в котором находим 46 вхождений в 44 документах.

Аналогично можно подвергнуть проверке *мне это надоело* и получить следующие результаты в InterCorp v8:

InterCorp v8 - Russian		InterCorp v8 - Polish	
<input type="checkbox"/> carroll-alenka_zrcadie	<p> - До чего мне это надоело ! - сказал вдруг после долгого молчания Шалтай-Болтай, не глядя на Алису.	carroll-alenka_zrcadie	<p> " To bardzo irytujące - rzekł po długim milczeniu Hojdy Bojdy, patrząc w kierunku przeciwnym niż Alicja - jak człowieka nazywają jakimś ... i to bardzo! " </p>
<input type="checkbox"/> _SUBTITLES	Мне это надоело .	_SUBTITLES	Przepraszam.
<input type="checkbox"/> _SUBTITLES	Знаешь, мне это надоело .	_SUBTITLES	Wiesz, nie muszę tu być.
<input type="checkbox"/> _SUBTITLES	Мне это надоело Хочешь подражаться?	_SUBTITLES	- Odwal się. - Chcesz się bić?
<input type="checkbox"/> Frankova-Denika_Franko	Как мне это надоело , не удивительно, что я испытываю к ней чуль ли не отвращение! </p>	Frankova-Denika_Franko	Uważam, że to okropne, ale powoli zaczynam nabierać do niej wstrętu! </p>
<input type="checkbox"/> _SUBTITLES	- Да что вы!, мне это надоело !	_SUBTITLES	- Dalej, jestem od tego chory!
<input type="checkbox"/> _SUBTITLES	Однажды мне это надоело , и я его вырубил.	_SUBTITLES	Pewnego dnia, nakarmił em go i ogłuszył em.
<input type="checkbox"/> _SUBTITLES	Нет, мне это надоело !	_SUBTITLES	Nie.
<input type="checkbox"/> _SUBTITLES	А потом мне это надоело , я сказал:	_SUBTITLES	Az powiedział em: " Pieprzyć to.
<input type="checkbox"/> _SUBTITLES	Мне это надоело !	_SUBTITLES	Zabijacie nas! Mam tego dosyć! Mam dość!
<input type="checkbox"/> _SUBTITLES	Мне это надоело .	_SUBTITLES	Mam dosyć.
<input type="checkbox"/> _SUBTITLES	Ну, всё, мне это надоело .	_SUBTITLES	Mam już tego dość.

Как свидетельствует корпусный материал (12 вхождений), и **мне это надоело** может составлять с фраземой ***mam tego dość*** пару эквивалентов.

Проверка может касаться также исходного польского транслянда, равно-значный вариант которого ***mam tego dosyć*** фиксирует корпус InterCorp v8:

InterCorp v8 - Polish		InterCorp v8 - Russian	
.SUBTITLES	- Mam tego dosyć .	.SUBTITLES	- Лучше бы он прекратил .
.SUBTITLES	Mam tego dosyć .	.SUBTITLES	Мне надоело вкалывать !
.SUBTITLES	Mam tego dosyć .	.SUBTITLES	Все, хватит ! Я пойду спать .
.SUBTITLES	Ja też mam tego dosyć .	.SUBTITLES	Это был Рутгерс , а не Браун поступив , я понял тогда , что если не займусь музыкой сейчас я не займусь ею никогда . Жаль , что у меня не было такой же четкой цели . Но надеюсь ... что скоро я ее найду .
Hasek-OsudyDobrehoVvSV	<p> - Tylko mi się tu nie bałwan , mój Szwajku , bo ja też mam tego dosyć . </p>	Hasek-OsudyDobrehoVvSV	<p> - Только , пожалуйста , не валяй дурака , Швейк ! Мне это уже надоело . </p>
rowlingova-hpot_vezen	- Mam tego dosyć !	rowlingova-hpot_vezen	- Мое терпение погнуло .
rowlingova-hpot_vezen	- Mam tego dosyć .	rowlingova-hpot_vezen	- С меня хватит , я в вашу игру не играю . </p>
.SUBTITLES	Mam tego dosyć .	.SUBTITLES	Я так устала .
.SUBTITLES	Mam tego dosyć !	.SUBTITLES	- Что ты такое говоришь ? - Надоело !
.SUBTITLES	- Mam tego dosyć !	.SUBTITLES	Я не прекращу борьбу .
.SUBTITLES	Mam tego dosyć !	.SUBTITLES	С меня довольно !
.SUBTITLES	Mam tego dosyć !	.SUBTITLES	Мне это надоело !
.SUBTITLES	Mam tego dosyć .	.SUBTITLES	Не понимаю .
.SUBTITLES	Mam tego dosyć !	.SUBTITLES	Хватит с меня !
hemingway-starec_a_mor	- Mam tego dosyć . </p>	hemingway-starec_a_mor	<p> - Ух , и надоел же ты мне со своими советами ! - сказал он вслух . </p>
.SUBTITLES	Mam tego dosyć .	.SUBTITLES	С меня хватит . Звоню в службу по защите детей !
.SUBTITLES	- Mam tego dosyć !	.SUBTITLES	Если ты с ним не справишься , отправь его в интернет . Хотя бы дома будет тише .
Hasek-OsudyDobrehoVvSV	<p> - Już mam tego dosyć ! - krzyknął kapral .	Hasek-OsudyDobrehoVvSV	<p> - Довольно с меня ! - вскричал капрал . - Я вас обоих в тюрьму могу упрятать . </p>
.SUBTITLES	Mam tego dosyć .	.SUBTITLES	Хватит с нас этого дерьма .

Следует обратить внимание на то, что фразема ***mam tego dosyć*** фиксируется в корпусе 19 раз, т. е. практически столько же, сколько ***mam tego dość*** (23). Это свидетельствует о том, что бо́льшая частотность употребления отдельного слова (***dość***), чем его тождественного по лексическому значению варианта (***dosyć***⁷), вовсе не означает бо́льшей частотности употребления фраземы, в которой оно употреблено.

Подробный анализ показывает очередной возможный переводной эквивалент — **с меня довольно**:

InterCorp v8 - Russian		InterCorp v8 - Polish	
<input type="checkbox"/> rowlingova-hpot_tajemn	Нет , с меня довольно !	rowlingova-hpot_tajemn	Mam już tego dosyć !
<input type="checkbox"/> .SUBTITLES	С меня довольно !	.SUBTITLES	Mam tego dosyć !
<input type="checkbox"/> .SUBTITLES	- Нет , с меня довольно !	.SUBTITLES	- Nie , dość tego !
<input type="checkbox"/> .SUBTITLES	С меня довольно .	.SUBTITLES	Tyle wystarczy . / Czy pomóc w czymś jeszcze ?
<input type="checkbox"/> .SUBTITLES	С меня довольно .	.SUBTITLES	Dość tego .
<input type="checkbox"/> .SUBTITLES	Ну все , с меня довольно .	.SUBTITLES	To ma sens !
<input type="checkbox"/> .SUBTITLES	С меня довольно .	.SUBTITLES	Mam już dość .
<input type="checkbox"/> .SUBTITLES	С меня довольно !	.SUBTITLES	- Kiedy wreszcie zrozumiesz , że nie musi kierować toba strach ?
<input type="checkbox"/> .SUBTITLES	С меня довольно .	.SUBTITLES	Brednie ! Nie chcę słuchać tych ohydnych plot .
<input type="checkbox"/> .SUBTITLES	С меня довольно .	.SUBTITLES	Mam już dość .
<input type="checkbox"/> .SUBTITLES	Безумие – Пожалуйста , послушай меня . - Нет , с меня довольно	.SUBTITLES	To fantazja . To szaleństwo . - Posłuchaj mnie , proszę . - Nie . Nie chcę słyszeć ani jednego

7 В польском языке ***dość*** употребляется чаще, чем ***dosyć*** (ср. BAŃKO, M. (red.): *Inny słownik języka polskiego*. Т. 1. Warszawa 2000, s. 308).

Следует заметить, что *с меня довольно* фиксируется в русско-польском корпусе InterCorp v8 лишь 11 раз. Эту статистическую информацию следует сопоставить с употребительностью других возможных эквивалентов, поскольку это позволяет определить очерёдность фиксации транслятов в словаре.

Как можно заметить, проверка русских эквивалентов не только может подтвердить правильность установленных эквивалентов, но и выявить варианты исходных транслятов. Так, корпус InterCorp v8 фиксирует 12 вхождений *mam już tego dość* и 7 *mam już tego dosyć*. Кажется, что и эти варианты заслуживают фиксации в переводном словаре. Поскольку количество вхождений обоих вариантов ограничено, частотность их употребления можно дополнительно проверить в Национальном корпусе польского языка (НКЯ): *mam tego dość* (в сбалансированном — 70, в полном — 199), *mam już tego dość* (соответственно: 28, 98), *mam tego dosyć* (соответственно: 44, 134), *mam już tego dosyć* (соответственно: 10, 61). Корпусные данные однозначно подтверждают воспроизводимость этих вариантов.

Как было уже отмечено, скрупулёзный анализ материала, предоставленного корпусом InterCorp v8, обнаружил и эквивалент *всему есть предел*, который заслуживает отдельного рассмотрения. Кажется, что в определённых контекстах его польским соответствием может быть *wszystko ma swoje granice* с возможными вариантами. Значение и ситуацию употребления этой фраземы приводит Иной словарь польского языка:

«mówimy „Wszystko ma swoje granice”, jeśli ktoś w swoim postępowaniu posunął się za daleko i oceniamy to negatywnie»⁸.

Wszystko ma swoje granice имеет 39 вхождений в сбалансированном и 180 в полном НКЯ.

Толковый словарь русского языка под редакцией Д. В. Дмитриева приводит следующее значение *всему есть предел*:

«Фраза «Всему есть предел» употребляется в том случае, если что-либо сказанное, сделанное кем-либо ощущается другим человеком как крайне неприличное, неуместное»^{9,10}.

Как можно видеть, в обоих языках фраземы *wszystko ma swoje granice* и *всему есть предел* употребляются как эмоциональная реакция для выражения негативной оценки сложившейся ситуации. Как правило, она связана с чьими-либо действиями или отсутствием таковых. Следует обратить внимание

8 BAŃKO, M. (red.): *Inny słownik języka polskiego*. T. 1. Warszawa 2000, s. 478.

9 DMITRIJEV, D. V. (red.): *Tolkovyj slovar' russkogo jazyka*. Moskva 2003, s. 982.

10 Стоит заметить, что исследуемые фраземы определяются в приведённых словарях с помощью прагматической дефиниции (см. GROCHOWSKI, M.: *Konwencje semantyczne a definiowanie wyrazów językowych*. Warszawa 1993, s. 34–39).

также на относительно часто предшествующие обоим фраземам противительные союзы *ale* и *no*. Это совсем не удивительно, так как в семантике противительности заложена одна цель — выражение субъективного отношения к действительности. Как *wszystko ma swoje granice* в польском, так и *всему есть предел* в русском языке имеют негативно-оценочный характер.

Корпус InterCorp v8 фиксирует также 1 вхождение *всему есть границы*, которое, казалось бы, также может соответствовать требованиям эквивалентности. По понятным причинам (лишь одно вхождение), следует обратиться к НКРЯ, в котором находим 8 вхождений фраземы *всему есть предел* в 8 документах. Анализ показывает, что она не употребляется в современном русском языке (первое употребление в 1842, последнее — в 1931 году). Эти корпусные данные имеют большое значение для лексикографа, поскольку исключают фиксацию этой фраземы в качестве переводного эквивалента. Как можно видеть на примере польской фраземы *tam tego dość*, использование параллельного корпуса решающим образом предопределяет возможности относительно установления её эквивалентов. Корпус выявляет не только варианты возможных транслятов, в том числе иных по лексическому составу, но и варианты самого исходного транслянда *tam tego dość*. В результате устанавливается не только одна пара эквивалентов, но извлекаются и другие, которые создают своеобразную картину польско-русской эквивалентности. Эта картина дополняется и синонимическими фраземами, которые выявляются в процессе поиска эквивалентов. Как было показано выше, этот поиск выходит за рамки параллельного корпуса. В нём используются и одноязычные корпуса имеющие большие размеры, а также осуществляется метод обратного перевода.

Несмотря на возможные ошибки, которые можно найти в параллельном корпусе, — он надёжный источник как трансляндов, так и транслятов. Сами субтитры, которые составляют в корпусе InterCorp v8 значительную долю, являются именно результатом процесса перевода с учётом всех компонентов вышеупомянутой «прагматической рамки». Немаловажно и то, что переводы осуществляется многими переводчиками, носителями разных идиолектов. Кроме того, они применяют разные переводческие стратегии, по-разному интерпретируют значение отдельных фразем исходного языка, наконец, находят разные эквиваленты в языке перевода. Благодаря этому можно избежать возможных ошибок, которые в процессе перевода могут быть сделаны.

Конечно, субтитрование как вид аудиовизуального перевода далеко не совершенно. Оно имеет и свои недостатки — иногда не отражает того, что говорится на экране. Субтитрованию не чужды компрессия и опущения, которые в известной степени могут терять отдельные компоненты речи, напр., её экспрессивность.

Факторы чисто технического характера, связанные со скоростью чтения зрителя, временем отображения субтитров на экране и сменой планов в кадре, навязывают, как правило, ограниченное количество строк и знаков. Всё это приводит к одному основному последствию — текста на экране меньше, чем содержат диалоги фильма. Эта особенность субтитров даёт основание говорить о них как о сокращённом переводе диалогов фильма. Эти недостатки отмечаются исследователями¹¹. Следует иметь в виду факт, что параллельный корпус может весьма нередко фиксировать перевод с английского на польский или с английского на русский. Следовательно, составляемые пары эквивалентов будут фактически польскими или русскими транслятами английского транслянда. Несмотря на эти недостатки, трудно не видеть огромные возможности, которые эти корпуса открывают для переводной лексикографии. Следует отметить, что субтитры в большой степени могут помочь фиксировать в переводных словарях те единицы, которые либо вообще в них не фиксировались, либо фиксировались в ограниченном масштабе — т.е. единицы общения. Как правило, субтитры отражают живую речь, которая предоставляет лексикографу очень ценный материал для дальнейшей обработки. Иными словами, с помощью субтитров получается уникальный материал, который можно сравнивать лишь с трудоёмкой записью устной речи.

Существенной технической проблемой параллельных текстов, которая полностью не решена, является автоматическая обработка текстов, а прежде всего автоматическое выравнивание текста-оригинала и текста-перевода на уровне слов, словосочетаний и предложений. По известным причинам процесс автоматического выравнивания устных текстов (высказываний) гораздо сложнее письменных (предложений).

Представляется, что параллельные корпуса текстов должны прежде всего учитывать то, что в языке типичное, закономерное, повторяющееся, воспроизводимое. Так, параллельные корпуса, содержащие разные тексты разных жанров — это идеальный инструмент для переводной лексикографии.

Литература:

BOGUSŁAWSKI, A.: *Podstawy konfrontatywnej lingwistyki przekładowej*. Łask 2013. ISBN 978-83-60178-07-1.

11 GORŠKOVA, V. Je.: *Osobnosti perevoda fil'mov s subtitrami*. Vestnik Sibirskogo gosudarstvennogo aërokosmičeskogo universiteta imeni akademika M. F. Rešetneva, Vyp. 3 (10), Krasnojarsk 2006, s. 142.

- GORŠKOVA, V. Je.: *Osobnosti preвода fil'mov s subtitrami*. Vestnik Sibirskogo gosudarstvennogo aërokozmičeskogo universiteta imeni akademika M. F. Rešetneva, Vyp. 3 (10), Krasnojarsk 2006, s. 141–144. ISSN 1816-9724.
- GROCHOWSKI, M.: *Konwencje semantyczne a definiowanie wyrażeń językowych*. Warszawa 1993. ISBN 83-85372-14-8.
- GRIŠINA, Je. A. – SAVČUK, S. O.: *Korpus ustnych tekstov v NKRJA: sostav i struktura*. In: Nacional'nyj korpus russkogo jazyka: 2006–2008. Novyje rezul'taty i perspektivy. Sankt-Peterburg 2009, s. 129–148. ISBN 978-5-98187-327-0.
- CHLEBDA, W.: „*Ramka pragmatyczna*” w procesie weryfikowania i tworzenia słownikowych definicji frazeologizmów. In: HORDY, M. – MOKIJENKO, W. – WALTER, H. (red.): *Język. Człowiek. Dyskurs*. Szczecin 2007, s. 185–195. ISBN 978-83-60903-36-0.
- KITAJGORODSKAJA, M. V. – ROZANOVA, N. N.: *Reč' moskvičej. Kommunikativno-kul'turologičeskij aspekt*. Moskva 1999. 5-93259-005-X.

Словари:

- BAŃKO, M. (red.): *Inny słownik języka polskiego*. T. 1. Warszawa 2000. ISBN 83-01-12826-7.
- DMITRIJEV, D. V. (red.): *Tolkovij slovar' russkogo jazyka*. Moskva 2003. ISBN 5-271-05995-2.

Корпусы:

- BAŃCZYK, Ł. – DYBALSKA, R. – VAVŘÍN, M.: *Korpus InterCorp — polština, verze 8 z 4. 6. 2015*. Ústav Českého národního korpusu FF UK, Praha 2015. <http://www.korpus.cz>. [online]. [cit. 11. 4. 2017].
- RAJNOCHOVÁ, N. – RUNŠTUKOVÁ, N. – VAVŘÍN, M.: *Korpus InterCorp — ruština, verze 8 z 4. 6. 2015*. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz>. [online]. [cit. 11. 4. 2017].
- NKJP: Narodowy Korpus Języka Polskiego, dostęp na stronie: <http://www.nkjp.pl>. [online]. [cit. 11. 4. 2017].
- NKRJa: Nacional'nyj korpus russkogo jazyka, režim dostupa: <http://www.ruscorpora.ru>. [online]. [cit. 11. 4. 2017].

Об авторе

Andrzej Charciarek, University of Silesia,
Faculty of Philology, Institute of East Slavonic Philology,
Katowice, Poland, andrzej.charciarek@us.edu.pl

